



OCR WEBINAR, 5 MARCH 2014, 1300 EET

Can crowdsourcing solutions serve many masters? Can they be beneficial for both, the laymen or native speakers of minority languages on the one hand and hard-core linguistic research on the other hand? How the infrastructure that was designed to support the linguistics, could turn out to be a solution that would raise the awareness of native languages? In the course of the **Digitization Project of Kindred Languages**, the **National Library of Finland** has developed an open source code OCR editor that offers some conceivable answers to the highlighted questions.

The OCR editor enables the editing of machine-encoded text for the benefit of linguistic research by crowdsourcing. The tool works on PDF files. It supports columnization, makes it possible to use very rare special characters used in the languages, and enables exporting the corrected text as required by the researchers. The user management features of the tool make it possible to distribute the correction work and its supervision.

The National Library of Finland is delighted to invite you to the OCR webinar, where the research tools and crowdsourcing solutions for editing the OCR text of digitized documents will be showcased for the public. The webinar will take place at the premises of the National Library of Finland on Wednesday, the 5th of March at 13:00 and online.

Time: 5 March 2014, 13-15 (EET); 12-14 (CET); 15-17 (MSK)

Place: Teollisuuskatu 23, Room A414, 4th Floor

Target group: The webinar aims to foster interdisciplinary work and linking together participant engaged in

- Text digitization and OCR
- Digital Humanities
- Digital Libraries
- Crowdsourcing
- Computing linguistics
- Speakers and researchers of minority languages

Programme: The webinar focuses on the topics related to editing the OCR. The following speakers will take the floor:

- Jussi-Pekka Hakkarainen (National Library of Finland): *An introduction to the OCR Webinar*
- Wouter van Hemel & Anis Moubarik (National Library of Finland): *What the OCR editor does and how it works*
- Jack Rueter (University of Helsinki): *The Language Access and Open Choice*
- TBC (Institute for the Languages of Finland): *The Kotus Experience: Crowdsourcing the Old Literary Finnish for the research's benefit*

The addresses will be followed by **a crowdsourcing session** in which the attendees may use the editor and frame questions regarding the OCR editor and working with it.

Registration form: <https://elomake.helsinki.fi/lomakkeet/49173/lomake.html>

Remote access to the Webinar (via Adobe Connect): https://connectpro.helsinki.fi/kk_finna/

Additional information and background material:

Basic information about the OCR editor: <http://blogs.helsinki.fi/fennougrica/2014/02/21/ocr-text-editor/>

Fenno-Ugrica Collection: <http://fennougrica.kansalliskirjasto.fi/>

Login Page for the OCR Editor: <http://ocrui-kk.lib.helsinki.fi/>

Contact address: kk-fennougrica@helsinki.fi

The webinar will be conducted in English.

Enrolling latest by the 3 MARCH 2014 at 1200 EET.